Calculating evapotranspiration: towards an operational Global Ecosystem Restoration Index (GERI)

MSc Thesis - Technical University of Denmark

Samuel Bloch

Abstract

The current decline in biodiversity prompted researchers to find indicators that would help them monitor and counter this trend. The Global Ecosystem Restoration Index (GERI) is such an indicator; it is dedicated to fight land degradation and restore 15% of degraded land by the end of the decade, a biodiversity target set in Aichi in 2010 by the members of the Convention on Biological Diversity. In this context, this study focuses on the computation of evapotranspiration to develop an indicator, the Evapotranspiration Ratio (ER), which would allow researchers to detect degradation hotspots in a given area, through the analysis of temporal trends. Based on the Penman-Monteith equation, the model was adapted from a previous study and improved; meteorological data from the global reanalysis ERA-Interim were included, and a workflow was designed so that future users would have few manual manipulations to do before obtaining their results. Validation was performed using field data from Coussoul, in the south of France; it returned mixed results, with meteorological data behaving very well while other variables showed rather bad correlation with physical measurements. In the end, simulations in the period 2003-2015 were performed and correlation between time and ER was calculated, with significant points ($p<0.05$, $r^2>0.8$) labelled as potentially degraded.

I. Introduction

Biodiversity is a contracted version of "biological diversity", defined by the Convention on Biological Diversity as "the variability among living organisms from all sources [...]; this includes diversity within species, between species, and of ecosystems." Therefore, it includes concepts such as genetic variety within a species, or habitat variety within a landscape (Encyclopedia of Earth, EoE (2016)). Today, biodiversity is declining at an alarming rate. According to the World Wildlife Fund (WWF, 2016), the loss of species we see today is 1000 to 10000 times higher than the natural extinction rate, meaning that between 0.1 and 0.01% of all species on Earth become extinct each year. This mass extinction event is often referred as the 6th mass extinction, but unlike the previous one, it is caused by one and only one species, the human .

In 1993, the United Nations Convention on Biological Diversity entered into force after being signed by 168 states. It has three targets (CBD, 2016):

1. The conservation of biological diversity
2. The sustainable use of the components of biological diversity
3. The fair and equitable sharing of the benefits arising out of the utilization of genetic resources

In 2010, at the Aichi conference (Japan), a "revised and updated Strategic Plan for Biodiversity" (CBD, 2016) was adopted for the 2011-2020 period, labelled as the UN Decade for Biodiversity. This plan set the 20 Aichi biodiversity targets , spread among 5 strategic goals that are to be reached by the end of the decade. For instance, Target 15 states: "By 2020, ecosystem resilience and the contribution

of biodiversity to carbon stocks has been enhanced, through conservation and restoration, including restoration of at least 15 per cent of degraded ecosystems, thereby contributing to climate change mitigation and adaptation and to combating desertification". GEO Biodiversity Observation Network (GEO-BON, 2016) is an international organization that works towards accomplishing these targets.
GEO BON is a consortium part of GEO, The Group on Earth Observations, and a partner of the CBD. GEO BON aims at collecting, studying and delivering biodiversity observations to a set of customers including decisions makers, NGOs and scientists. In relation to the Aichi biodiversity targets, the organizations has defined Essential Biodiversity Variables (EBV), "window[s] into the biodiversity observation systems" to be used by its customers . GEO BON's aim is to produce indicators "characterized by the rigorous use of open access large global datasets, state of the art remote-sensing based information, model-based integration of multiple data sources and types, including in situ (ground based) observations, and online infrastructure enabling inexpensive and dynamic updates, with full transparency" (Fernandez et al., 2015), which assess progress towards the Aichi targets. One of these indicators is the Global Ecosystem Restoration Index (GERI, Fernandez et al. (2015)), aimed at Target 15 (restoration of 15% of degraded ecosystems). The GERI is made of three components related with ecosystem health and degradation status: changes in land productivity, changes in surface energy balance and changes in land cover. "This composite index is produced for near the entire terrestrial surface of the planet at a spatial resolution of 1 km2. This allows for the GERI index to be aggregated to small regions, states, countries, continents and the planet. This index uses well vetted products derived from MODIS sensor, in orbit since 2001. These products are continuously being outputted and the expected data availability and planned mission continuity is an assurance of the temporal sustainability of the proposed index." The GERI addresses the two following essential biodiversity variables: net primary productivity and ecosystem extent and fragmentation.

Researchers can now count on remote-sensing to monitor biodiversity. In the past, information regarding biodiversity was gathered in the field by making species census, for instance nest census. This kind of information was local and non-systematic or comparable across different sites or periods of time. In order to account for biodiversity losses, there is an urgent need to have systematic and comparable indicators of biodiversity across the globe and consistent over time; that is what the GEO BON program intends to do. In this sense, technological improvements made satellite data research relevant for this purpose; and linking biodiversity to remote sensing has become an important research topic in the past decade (Turner et al., 2003).
Two general models of biodiversity monitoring using remote sensing exist: direct and indirect models. The direct methods track "individual organisms, species assemblages or ecological communities", whereas the indirect ones rely on proxies, such as type of habitat (e.g. woodland, drylands) to estimate species range or species richness (Turner et al., 2003). The indirect methods are based on the assumption that there is a relationship between the proxy and the species richness, e.g. an old, pristine rainforest is expected to shelter more diverse forms of life than a soybean plantation at the same location.

Due to the complexity of biodiversity, combining different environmental parameters, quantitative and qualitative, is necessary to obtain better assessments of biodiversity. Thus, regarding the indirect method, Turner et al. (2003) identified three key indirect variables: "primary productivity, climate and habitat structure (including topography)". Duro et al. (2007) chose four indicators that can be derived from remote-sensing data: productivity, disturbance, topography and land cover. Land cover is a class-based indicator, unlike others that

are based on continuous mapping. That means that when using land-cover maps based on remote-sensed data and in-situ data, the Earth surface is divided into classes or blocks, each representing a certain category of land-cover (generally vegetation). Pettorelli et al. (2016) pointed out that such classification needs agreement among researchers, an agreement that is apparently difficult to reach .
Two other issues related to mapping in ecological research are pointed out by Fassnacht et al. (2006): scale (the combination of extent, grain and minimum mapping unit of a study) and accuracy assessment (e.g. how error is distributed on a map, and how it affects further analysis).

This thesis is a contribution to the surface energy balance indicator of the GERI, to address degradation and restoration of ecosystems. To use land degradation as a proxy for biodiversity, it is assumed that the more degraded the ecosystem, the lower the biodiversity (Fernández, 2013).
Then, the problem of using indirect indicators to monitor biodiversity in our case starts with the problem of developing a land degradation indicator.
Land degradation needs to be defined. In 1996, the United Nations Conventions to Combat Desertification (UNCCD) defined it as a "reduction or loss in the biological and economic productivity and complexity of terrestrial ecosystems, as well as in the ecological, biochemical and hydrological processes that operate in them" (UNCCD, 1996).

In general, most of the remote sensing indicators of land degradation try to detect changes in the relationship between plant Net Primary Production (NPP) and rainfall. The Rainfall Use Efficiency index (RUE) uses the ratio of NPP to precipitation to explain land degradation; it's intended to separate climatic effects (e.g. changes in annual rainfall) from human induced land degradation (e.g. land cover changes). Prince et al. (1998) noted that this index had been locally applied in several areas, however, indices like the RUE or others relying heavily on the Normalized Difference Vegetation Index (NDVI) "'tend only to indicate susceptibility to desertification rather than actual degradation"'. Moreover, Huete and Jackson (1987) noted that a "non-linear scattering of photons" from a bright arid soil could mask reflectance signals from vegetation (especially senesced grass), hence the questioning of vegetation-based indicators in arid environments.

García et al. (2008) showed that a new type of land degradation indicator related to surface water deficit, for instance the NEFS (Non-Evaporative Fraction Standardized) could successfully lead to the detection of not only extremely disturbed or undisturbed areas in southern Spain but also areas that had suffered more subtle disturbances (e.g. topsoil losses). This study suggests that such indicators, derived from satellite datasets, from ASTER and MODIS sensors for instance, can give an insight into land degradation and then be connected to biodiversity losses beyond more classic remote sensing indicators of degradation based on analysis of vegetation greenness García et al. (2008). "Functional ecosystems should optimize the fraction of dissipated energy in the land surface in the form of [evapotranspiration]. Therefore the partition of the available energy reaching the land surface into latent heat [evapotranspiration] and sensitive heat provides key information on the restoration process" (Fernandez et al., 2015).

To estimate an energy-based indicator such as the NEFS from remote sensing information, it is necessary to calculate evapotranspiration and net radiation fluxes. Evapotranspiration is the rate of water loss from an ecosystem to the atmosphere, including soil and water bodies (evaporation) and plants, through stomata in their leaves (transpiration). The calculation of the indicators related with the evaporative fraction (or its reciprocal non-evaporative fraction) requires the estimate of evapotranspiration. This topic has been the subject of many studies

(Kalma et al., 2008) using several models and remote sensing data. They are all based on the surface energy balance equation:

$$Rn - G - H - LE = 0 \quad (1)$$

where Rn is the net radiation, G the ground heat, H the sensible heat and LE the latent heat, i.e. the expression of evapotranspiration in terms of energy.

We consider broadly two types of methods to estimate LE relying on remote sensing datasets. A first group of models contains those which calculate LE as a residual of the energy balance equation (Kalma et al., 2008). They are based on the difference between surface temperature and air temperature, which drives H, and require to calculate the aerodynamic resistance to H, turbulent heat transfer. It is therefore H that is solved first, LE being the residual. Two-source residual models, treating land surface as a composite of soil and canopy, have a "strong physical basis" and provide better results in drylands than one-source models (Anderson et al., 2007), but they are difficult to apply regionally because the estimation of aerodynamic resistance requires precise knowledge about parameters related with vegetation roughness or atmospheric stability.

A second type of models directly solves for LE. The Penman-Monteith combination equation is the basic equation of the second group of models that use remote-sensing. It combines the surface energy balance and the mass-transfer flux equations for latent and sensible heat fluxes. Here, when applying this method globally or regionally, the challenge is to evaluate the variations of surface conductance to water vapor (García et al., 2013) without using field calibration. Several methods have been proposed, with constant parameters throughout the year (Leuning et al., 2008) or with variable parameters depending on rainfall (Zhang et al., 2008).

Priestley and Taylor suggested in 1972 a simplification of the Penman-Monteith equation that got rid of the surface and aerodynamic resistances to replace them by an empirical multiplier, α. That way, the model doesn't require explicit estimates of water vapor deficit or stomatal conductance, although it may be less accurate than a Penman-Monteith model and is hard to apply for non-potential conditions. Fisher et al. (2008) proposed a model (PT-JPL) based on the Priestley-Taylor equation to calculate monthly the actual evapotranspiration using empirical multipliers derived from remote sensing and meteorological datasets (biophysical constraints) to the potential evapotranspiration to obtain actual evapotranspiration.

Because they are based on remotely-sensed data, all these models face the problem of data consistency, especially regarding time scaling (e.g. MODIS data may be available daily, or every 8 days, or every 16 days) and space scaling (e.g. data may be available with pixels representing 1 km, or 10 km).

The PT-JPL model has shown great results at monthly time scales in many kinds of environments, "ranging from boreal to temperate and tropical ecosystems" (García et al., 2013), but not for drylands receiving less than 400mm of rain every year (Fisher et al., 2008)(Fisher et al., 2009). Given that such habitats are particularly sensitive to water availability changes and land degradation, García et al. (2013) adapted the PT-JPL model to make it work on a daily time scale, with a new way to calculate soil moisture based on the thermal inertia concept (Verstraeten et al., 2006). Based on data from two sites in Sahel and Spain, they showed that this rather simple model gave equivalent results whether using data from remote sensing or field data in Sahel. The results from the site in Spain were less accurate, but were still better than those from other models (a residual two source energy balance model from Kustas and Norman (1999) and Penman-Monteith-Leuning) run in the area (Morillas et al., 2013b) (Morillas et al., 2013a). Expanding the PT-JPL model, tested at the point level in two flux sites in García et al. (2013), to apply it to a regional level and using only freely available EO (Earth Observation) data is a necessary step to develop a global degradation index,

involving technical and conceptual challenges.

II. Objectives

The general aim of this thesis was to contribute to the further development of the GERI indicator at regional level as a proof of concept. In particular, the study focused on the GERI component that tracks changes in surface energy balance and evapotranspiration using a remote sensing –based model. Three particular objectives have been addressed: script improvement, validation and detection of degradation hotspots. The starting point of this thesis was the Matlab script used by García et al. (2013) and a preliminary regional version from Moyano et al. (2015) for their studies. From a very specific script, customized for one user, one computer and one region, we tried to offer a generic framework in which the user doesn't need many manual operations to produce results. In addition to that, meteorological data from global reanalysis were introduced to the model, which then needed to be validated. The final objective was to analyze temporal trends to detect degradation hotspots in the study area.

Figure 1: *General location of the area of interest (orange square)*

III. Data and study area

I. Study site: Coussoul and the south-east of France

The area of interest is located in the south-east of France. This choice was motivated by two reasons: initially, we wanted a country for which biodiversity data was available. A few of them were available, but as the focus of García et al. (2013) was drylands, we chose an arid zone. Figure 1 shows the general location of this area. It was selected as a pilot site to test the functioning and accuracy of the spatial model. In subsequent stages, after this project, the model can be applied in a larger region. The pilot area, or study area, covers different land use types, including wetlands with high biodiversity (Camargue) and flux sites for validation.

II. Validation data

For validation purposes, Sébastien Garrigues (INRA-EMMAH, Avignon, France) provided field data from Coussoul, a dry grassland site in the Crau-Camargue area (43.56 N; 4.86 E, (Mira et al., 2016)). Measurements started on this site in 2010, with a CNR1 net radiometer (Kipp Zonen, Delft, The Netherlands) that measures "incoming and outgoing irradiances in the solar domain (global radiation, or shortwave radiation) and the thermal infrared domain (atmospheric radiation, or longwave radiation)" (eddy covariance system). Albedo was then calculated as the ratio of outgoing shortwave radiation over incoming shortwave radiation, and "surface temperature was computed from the outgoing thermal irradiance (LWo) based on the Stefan–Boltzmann law". Instruments were placed between 1.5 and 2 meters above the ground, and "measurements were made every second and averaged every 30 min".

III. MODIS data

MODIS (Moderate resolution imaging spectroradiometer) is an instrument aboard two NASA

satellites, Terra and Aqua, which acquires data in 36 spectral bands or group of wavelengths. Orbiting the earth, these two satellites allow NASA to provide various datasets (MODIS products) that can be used freely by scientists. Table 1 shows the variables we use to calculate evapotranspiration, classified by the MODIS products they are extracted from. While these data are all available with the same spatial resolution (1 km), the time steps between two values were not always the same. Therefore, when data were missing, measurements from the last available date were used.

MODIS data are available in tiles, each one covering a certain part of the planet. Our area of interest was entirely located on tile h18v4, therefore data were downloaded for this tile only.

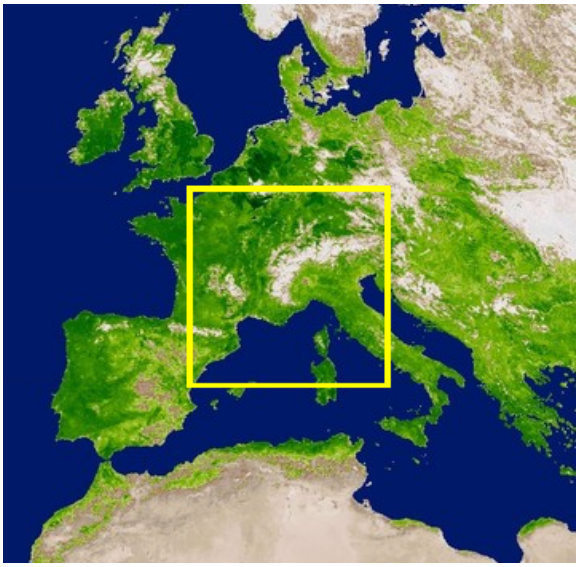

Figure 2: *Tile h18v4. The contour is shown in yellow*

Data from the Terra satellite are available from 1999, but measurements from Aqua start in the middle of 2002. Thus, our analysis covers the period 2003-2015.

IV. ERA-Interim data

ERA-Interim is a daily global atmospheric reanalysis offering data from 1979 to now, updated every month with a delay of two months. Among the many meteorological parameters available in the dataset, we were interested in two of them: air temperature at 2 meters, and surface solar radiation downwards.
Some parameters are called 'analysis parameters'. Available four times a day, at 0, 6, 12 and 18 UTC, they mostly rely on observation. From these parameters, 'forecast parameters' are calculated. Unlike the previous ones, they mostly rely on the ERA model, not on observation. Still, they are based on an initialization epoch (or analysis time, 0 or 12 UTC), and are available 3, 6, 9 or 12 hours after this epoch (not at the initialization time).
Given these constraints, we downloaded the required data with the following settings, so that they would be as close as possible to the time when the satellite passed overhead (which is around 11 in the study area):
- 2 meter temperature is an analysis parameter, so we downloaded it at time 12, step 0.
- Surface solar radiation downwards is a forecast parameter, so we downloaded it at time 0, step 12.

The spatial resolution of the ERA dataset is approximately 80 km, whereas the MODIS resolution is 1 km.

Figure 3 provides an overview of the area of interest, including the ERA data points, the partitioning related to them and the validation site of Coussoul.

IV. Methods

I. The PT-JPL model

The model we run is based on the study from García et al. (2013) and modified for a preliminary spatial version by Moyano et al. (2015). It is a daily implementation of the PT-JPL model, which is based on the algorithm described by Fisher et al. (2008). In this paper, the word "evapotranspiration" refers to latent heat, the energy flux equivalent of the physical process called evapotranspiration. "Real" evapotranspiration is expressed in grams or millimeters, but latent heat is expressed in $\mathrm{W\,m^{-2}}$. There-

Table 1: *MODIS products used in this study. NDVI means Normalized Difference Vegetation Index, fPAR means Fraction of Photosynthetically Active Radiation, LAI means Leaf Area Index.*

MODIS product	Time step	Variable
MOD13A2	16 days	1 km NDVI
MOD15A2	8 days	1 km fPAR
	8 days	1 km LAI
MCD43B3	8 days	1 km Black/White sky albedo
MOD/MYD 11A1	1 day	1 km Band 31/32 Emissivity
		1 km Daytime land surface temperature
		1 km Nighttime land surface temperature
		1 km Day observation time
MOD/MYD 11A2	8 days	1 km Band 31/32 Emissivity
		1 km Daytime land surface temperature
		1 km Nighttime land surface temperature

fore, what we call evapotranspiration is expressed in $\mathrm{W\,m^{-2}}$ in this paper.

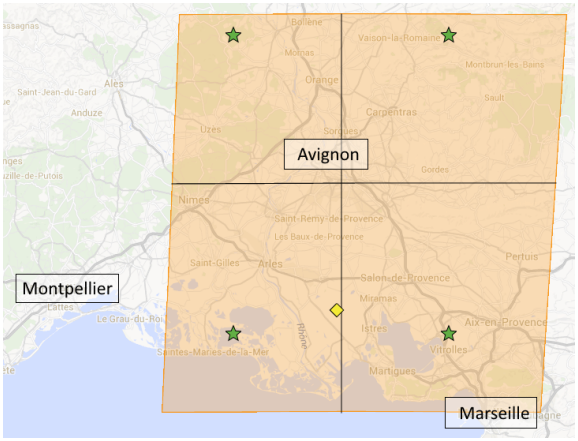

Figure 3: *Details of the area of interest. The 4 stars represent the locations of the ERA data points. The lines define 4 regions, one for each ERA data points. Pixels in each region share meteorological data from the ERA data point located inside this region. The diamond shape locates the validation site, Coussoul. The city and town names of Marseille, Montpellier and Avignon are also displayed.*

Actual evapotranspiration (LE) is calculated as a fraction of potential evapotranspiration (LEp), which is derived from net radiation (Rn) received by the habitat. Rn is made of two components, longwave and shortwave radiations, which in turn are made of two components, incoming and outgoing longwave/shortwave radiations, as described in Figure 4.

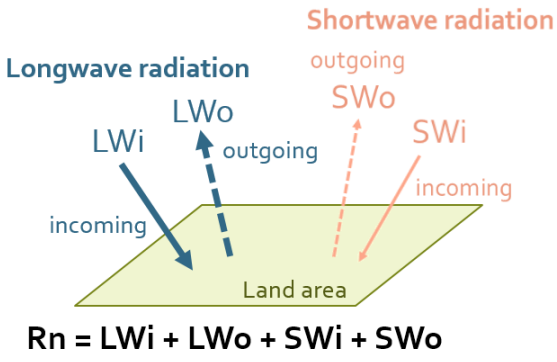

Figure 4: *Net radiation*

Longwave radiation (in $\mathrm{W\,m^{-2}}$) is computed as explained in Equation 2 and Equation 3 (Stefan-Boltzmann law), where T is the land surface temperature (K) and Emissivity the average emissivity of bands 31 and 32 (no unit). 5.67×10^{-8} is the value of the Stefan-Boltzmann constant, in $\mathrm{W\,m^{-2}\,K^{-4}}$.

$$LWi = 5.67 \times 10^{-8} * T^4 * 0.261 e^{-7.77 \times 10^{-4} T^2} \quad (2)$$

$$LWo = -5.67 \times 10^{-8} * \text{Emissivity} * T^4 \quad (3)$$

Incoming shortwave radiation (SWi) is provided by the ERA-Interim dataset. Outgoing shortwave radiation (in $\mathrm{W\,m^{-2}}$) is a fraction of SWi, as shown in Equation 4. Albedo, or reflection coefficient (no unit), is the broadband reflectivity of the land.

$$SWo = \text{Albedo} * SWi \tag{4}$$

The PT-JPL model is a two source model separating soil and vegetation fluxes. Therefore, actual and potential evapotranspiration, like net radiation, are partitioned into canopy evapotranspiration, occurring in the vegetation, and soil evapotranspiration, occurring on the ground (all in $\mathrm{W\,m^{-2}}$). Table 2 defines several notations for these variables, which are then defined in the subsequent equations.

Table 2: *Soil and canopy radiations - notations. ET means evapotranspiration. LE is called "actual evapotranspiration", or simply "evapotranspiration". All this variables are expressed in* $\mathrm{W\,m^{-2}}$.

Notation	Variable
Rn	Net radiation
Rn_c	Canopy net radiation
Rn_s	Soil net radiation
LE	(Actual) Evapotranspiration
LE_c	Canopy evapotranspiration
LE_s	Soil evapotranspiration
LEp	Potential evapotranspiration
LEp_c	Canopy potential ET
LEp_s	Soil potential ET

A layer model (additive fluxes) is assumed as follows:

$$Rn = Rn_c + Rn_s \tag{5}$$

$$LE = LE_c + LE_s \tag{6}$$

$$LEp = LEp_c + LEp_s \tag{7}$$

In the following equations, LAI is the Leaf Area Index, s is the slope of the vapour pressure curve (Pa), ψ is the psychometric constant (Pa), G is the ground heat ($\mathrm{W\,m^{-2}}$).
In order to calculate how much of the net radiation reaches the soil surface, a Beer Lambert law of light extinction is used with an extinction coefficient of radiation k=0.6 (Equation 8).

$$Rn_s = Rn * e^{-0.6*\mathrm{LAI}} \tag{8}$$

Potential evapotranspiration is calculated using the Priestley-Taylor (Priestley, 1972) approach:

$$LEp_c = 1.26 * \frac{s}{s+\psi} * Rn_c \tag{9}$$

$$LEp_s = 1.26 * \frac{s}{s+\psi} * (Rn_s - G) \tag{10}$$

To calculate actual evapotranspiration, we still need the multipliers to apply to potential evapotranspiration. These biophysical constraints, or stress multipliers, are similar to stomatal and soil conductance in the approach of Jarvis (1976) , and are defined in Table 3. The green canopy fraction (f_g) is the ratio between intercepted and absorbed photosynthetically active radiation. The plant temperature constraint (f_T) "accounts for reductions in photosynthetic efficiency when plants grow at temperatures departing from their optimum temperature range" (Potter et al., 1993). The plant moisture constraint (f_M) reflects changes in light absorptance with respect to the maximum $f_{APARmax}$. The soil moisture constraint (f_{SM-ATI}) is based on the thermal inertia concept, using land surface temperature and albedo. "Thermal inertia is a physical property of soil at the land surface measuring the thermal response of a material to the changes in its temperature" (Nearing et al., 2012).

The model is then complete with Equation 11 and Equation 12:

$$LE_c = f_g * f_T * f_M * LEp_c \tag{11}$$

$$LE_s = f_{SM-ATI} * LEp_s \tag{12}$$

In the end, we define the Evapotranspiration Ratio (ER) as in Equation 13. This is a

Table 3: *Biophysical constraints. NDVI means Normalized Difference Vegetation Index, PAR means Photosynthetically Active Radiation, ATI means observed Apparent Thermal Inertia index ($°C^{-1}$) (García et al., 2013)*

Name	Definition
Fraction of intercepted PAR	$f_{IPAR} = \text{NDVI} - 0.05$
Fraction of absorbed PAR	f_{APAR} (MODIS data)
Green canopy fraction	$f_g = \frac{f_{APAR}}{f_{IPAR}}$
Plant temperature constraint	$f_T = 1.1814 * \frac{1}{1+e^{0.3*(-35-T)}} * \frac{1}{1+e^{0.2*(15-T)}}$
Plant moisture constraint	$f_M = \frac{f_{APAR}}{f_{APARmax}}$
Soil moisture constraint	$f_{SM-ATI} = \frac{ATI-ATI_{min}}{ATI_{max}-ATI_{min}}$
Fractional vegetation cover	$f_c = f_{IPAR}$

dryness index that will be used in this study as an indicator for land degradation.

$$ER = \frac{\text{Evapotranspiration}}{\text{Potential evapotranspiration}} \quad (13)$$

II. Script improvement

Script improvement was performed with a focus on five different axes in order to improve the user's workflow, by minimizing the number of manual operations and making the code more user-friendly.
The five axes are the following:

1. Incorporation of meteorological data from a reanalysis dataset
2. Improvement of file reading in the script
3. Improvement of the definition of the spatial extent of the study area
4. Definition of a file structure
5. Creation of a tutorial for future users

The **incorporation of meteorological data from a reanalysis dataset** was necessary because, initially, the model used one single set of meteorological data (temperature at 2 meters and downwards surface solar radiation, or incoming shortwave radiation) for the whole area of interest. This was possible since the area considered in previous studies was small enough. With the introduction of data from the ERA-Interim dataset, a higher accuracy was expected at the regional scale we targeted. After the extraction of ERA data, each pixel in the study area is assigned a set of meteorological data. At the moment, this process is not fully automatized: based on the study area and the data points available in the ERA dataset, the user has to assign meteorological data to every pixel in the area.
The **improvement of file reading** was an important step for future users to be able to use the code with ease. The program reads many files during its execution, and therefore needs to be given the path leading to these files. Originally, this was done with absolute paths; we introduced relative paths.
Another interesting point is the **definition of the spatial extent of the study**. We are only studying a small part of the tile we download (tile h18v4 covers a large part of France but also the north-east of Spain, the south of Germany, Switzerland and the north of Italy), that's why we need to feed the script with the coordinates of our study area. Originally, these coordinates were repeated in the script every time they were used. Thus, when the user wanted to study a new area, he or she had to modify every occurence of these coordinates, and make sure that none was forgotten. This was a tedious and boring task, which could easily lead to mistakes. Instead of that, a generic notation for the spatial extent, with variable instead of numbers, was introduced.
This improvement process led to the **definition of a file structure** for the model, which can be downloaded by anyone wanting to run simula-

tions with it.
In the end a **tutorial for future users** was written, so that it's easy for them to run the model. The idea was to describe the file structure we created, and then detail the workflow that was defined so that anyone could run simulations, even with a limited knowledge of the code itself. This tutorial appears in the Appendix of this report.

III. Drought indicator validation

Validation of the model was performed using the results from year 2011, as advised by Sébastien Garrigues because it should be more reliable than other years. Data from the model were extracted from the results, for the place of measurements (Coussoul site, 43.6 °N 4.86 °E). To check the agreement between in situ datasets and EO (Earth Observations) or modeled datasets, we calculated three indicators: the Pearson product-moment correlation coefficient R, the root-mean-square error (RMSE) and the mean absolute error (MAE) for each of the following variables: net radiation (Rn), incoming (LWi) and outgoing (LWo) longwave radiation, incoming (SWi) and outgoing (SWo) shortwave radiation, evapotranspiration, albedo, air temperature and Evapotranspiration Ratio (ER).
The Pearson product-moment correlation coefficient (Pearson's r) measures the linear correlation between two variables, giving a result between -1 and 1. A value of 1 indicates total correlation, a value of -1 indicates total negative correlation, while a value of 0 indicates no correlation.

$$r = \frac{n\sum xy - (\sum x \sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (14)$$

Equation 14 defines Pearson's r, with x and y being realizations of two variables X and Y, and n the number of samples. In our case, X and Y represent the same physical variable, X being the output of the model (or input, for some variables like albedo and air temperature) and Y the field measurement of the same variable. n is the number of days on which comparison can be made, i.e. the number of days where both X and Y data are available (both datasets generally show gaps in data).
The root-mean-square error (RMSE) of a variable measures the difference between a prediction and the values actually observed. Equation 15 defines it, with x the predicted value of the variable, y the observed value, and n the number of data points available for comparison.

$$RMSE = \sqrt{\frac{\sum (x-y)^2}{n}} \quad (15)$$

The mean absolute error (MAE) also measures how close a predictor is to the variable it estimates. It is defined as the statistical mean of the absolute difference between predictor and observation, as shown in Equation 16.

$$MAE = \frac{1}{n}\sum |x-y| \quad (16)$$

While the Pearson's r explains how much the evolution of one variable explains the evolution of another variable, the RMSE and the MAE give an insight into the absolute error between those variables. Bias, as defined in Equation 17, can tell whether a certain error is random or systematic.

$$Bias = \frac{1}{n}\sum (x-y) \quad (17)$$

A bias of 0 indicates that the results of the model are unbiased. If the error is consistent (or systematic), the bias should be quite close to the MAE.
To evaluate the field flux data, we calculated the closure error CE%, as defined (in percent) in Equation 18. Rn means net radiation, G means ground heat, H means sensible heat and LE means evapotranspiration.

$$CE\% = \frac{Rn - G}{H + LE} * 100 \quad (18)$$

Ideally, the closure error should be 100%, meaning that the surface energy balance equation (Equation 1) is true.

IV. Detection of degradation hotspots

Once the performances of the model were evaluated, it was used for the detection of sites that might have been degraded or restored in the study area. This was achieved by calculating the correlation coefficient (Pearson's r or r-value, see above) between time and Evapotranspiration Ratio (ER). This value tells about how much one variable (time) explains the variations of the other (ER): once we find a place with a high correlation between the two, the next task is to look for the causes of this phenomenon, but this is out of the scope of the present study. In addition to the r-value, the p-value was also calculated for every pixel of the study area. It explains how significant the associated r-value is. With this information, we were able to consider only the places with high correlation between time and ER and with high significance.
MODIS data were available starting from 2003, so the analysis covered the period 2003-2015. A simulation was run for every odd year, resulting in the simulation of 7 years. The last year provided aberrant results, so it was discarded. For each pixel in the area, ER was averaged on a yearly basis, giving one value for each year. We then computed r and p-values based on this time-series. The square of the r-value, r^2, was scrutinized: the highest values highlighted the suspected degradation hotspots. Of these high r^2, we only kept those with a corresponding p-value inferior to 0.05, as suggested by the Matlab help.

V. Results and discussion

I. Script improvement

The steps taken to improve the script made the user's work considerably smoother.
The **incorporation of reanalaysis data** for some meteorological input variables proved successful: the correlation between them and field values was satisfying, as was the error (see the next section). However, the ERA-Interim dataset has a lower resolution than the rest of our input data, so we need to decide how these data are used. More specifically, we need to divide the study area into regions in which every pixel is assigned the same meteorological data. In our study, we used 4 ERA data points, and divided our study area into 4 regions in the shape of a rectangle. In each region, pixels were assigned meteorological data from the closest ERA data point, the one situated in the region. This process is not fully automatized yet: the user has to define these regions by itself. In the future, we hope that this partition can be done automatically, based on the definition of the study area.
Figure 5 shows how **file reading in the script** was improved.

```
% BEFORE (Absolute path)
read('F:/Documents/PT-JPL model/Spatial script/Data/MOD11A1/Tair.txt')
% AFTER (Relative path)
read('Data/MOD11A1/Tair.txt')
```

Figure 5: *File reading improvement in the script - from absolute to relative paths*

An absolute path (BEFORE) includes every folder from the hard drive letter to the file, whereas a relative path (AFTER) only includes the path from the working directory (in which the script is working, in this case 'Spatial script') to the file. That makes a significant difference because every computer is not structured the same way. For instance, hard drives can be assigned different letters (typically C, D, E...). To make the code usable by every potential user, we needed to get rid of this, lest the user has to modify the code to make it fit the file structure of his or her own computer.
The next step was the **definition of the extent of the study area**, with the use of variables instead of numbers throughout the script (see Figure 6) Now, when it starts, the main script calls a secondary script called 'parameters.m', in which the spatial extent of the study area is defined. Thus, the user needs to specify this extent of the study area in one and only one place, the 'parameters.m' script.

In the future, this script could also be used to define other variables that are repeated in the main script, if needed.

```
% BEFORE
Emis31_A=Emis31_A(595:636,344:385);
% AFTER
Emis31_A=Emis31_A(x1:x2,y1:y2);
```

Figure 6: *Spatial extent - from numbers to variables*

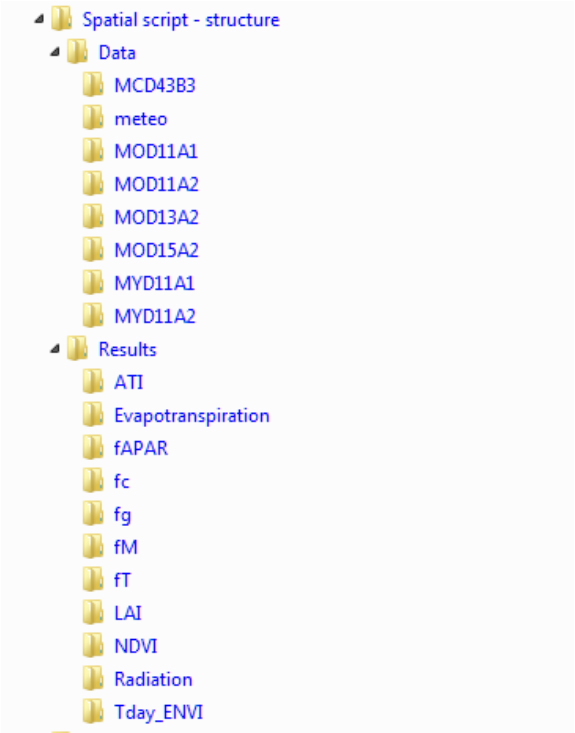

Figure 7: *File structure of the model*

In this project, we defined a **file structure** for the model; potential users can download it freely and adapt it to their needs. This framework, described in more details in the Appendix and shown in Figure 7, contains the scripts and text files needed for running a simulation but also the folders that contain input and output data. With this, the user knows where to save the data he or she downloads, and knows where the results of the simulation can be found.
In a previous version of the model, text files (corresponding to input data files) were generated manually by the user, based on these input files' names. The file structure we offer already contains most of the necessary text files; this results in a real gain of time for the user, but also affects his or her mood by removing boring and repetitive tasks.

Finally, a **tutorial** aimed at future users was written; it can be found in the Appendix. This tutorial explains the file structure and precisely describes the user's workflow. It also provides a list of input and output variables. We hope that it helps users understand how the scripts work together and what they need to do to run a simulation.

II. Drought indicator validation

To validate the results of our model and the remote sensing variables we use, we compared them to a field dataset. We assessed the validity of this dataset by calculating the closure error (Equation 18). Figure 8 shows that this error, which should ideally be around 100%, is particularly high. Even though we use it for validation, results for evapotranspiration and evapotranspiration ratio should be considered with caution due to this uncertainty.

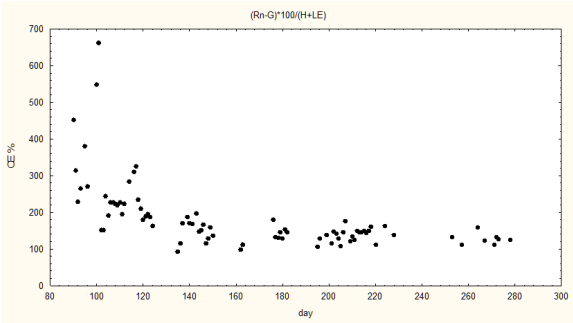

Figure 8: *Year 2011 - Closure error CE% (in percent) against time (in days)*

When data from the model or from MODIS are unavailable, they appear as NaN (not a number). However, unavailable field data appear as -9999, while several data points are also aberrant in the sense that they are different from -9999, but still odd ($-4000\,\mathrm{W\,m^{-2}}$ for sensible heat for instance). The first attempt at validation took into account every single field data point.
We could see that data from the ERA-Interim

dataset, air temperature and downwards surface solar radiation (SWi), were extremely well correlated to field data. Calculated longwave radiation was also relatively well correlated to the observations, but showed a consistent gap, or offset, between observation and prediction. We suspected that there was a systematic bias involved. Indeed, bias is 53.5 $\mathrm{W\,m^{-2}}$, i.e. 91% of the MAE. On the other hand, we noticed that outgoing short wave radiations gave badly correlated results.

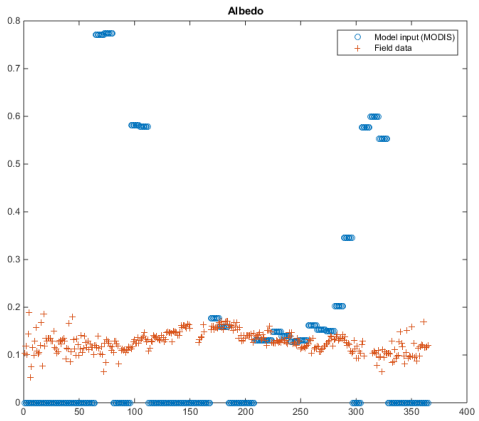

Figure 9: *Year 2011 - Albedo values from MODIS (circles) and the validation field dataset (crosses), against time (in days) - Area: Coussoul*

Given that $SWo = albedo * SWi$, we investigated albedo values (which we calculate as $Albedo = 0.8 *$ Black Sky Albedo $+ 0.2 *$ White Sky Albedo). We noticed that these values were really far away from the field data (Figure 9), in addition to missing data. High values (above 0.4) correspond to cloudy days, hence the high absolute values for outgoing shortwave radiation at the same time (not shown). Given that albedo is not calculated but downloaded from the MODIS database (product MCD43B3), we couldn't really conclude on the validity of our results.

To improve this aspect of the simulation, we decided to use a fixed value for the albedo, and apply it to the whole region. The chosen value was the mean of the albedo field data, 0.13. We expected this to improve our results, especially SWo. Doing so, we could see that, although the correlation between predicted and measured outgoing shortwave radiation was greatly improved, evapotranspiration and ER were not really improved (results not shown). We suspected that these results may be caused by aberrant data from the field measurements

Based on these observations, we filtered the aberrant field data: half-hourly values for evapotranspiration (LE), net radiation (Rn) and ground heat (H) were to be situated between -50 and 1500 $\mathrm{W\,m^{-2}}$ to be taken into account and aggregated into daily values. Further filtering steps dealt with the number of available measurements each day (one day wouldn't be taken into account if less than 80% of the half-hourly points were available that day) and the closure error ($CE = Rn - H - LE - G$, which should be inferior to LE), but these approaches didn't leave enough points for the comparison to be significant. Table 4, Figure 10 and Figure 11 show the result of the first filtering described above. Evapotranspiration correlation and error are improved, as is ER. We should still keep in mind, though, that field data are unreliable.

Subsequently, the simulations for the other years were run with a fixed value for the albedo, the same (0.13) for the whole area. It is also interesting to note some limits of our model: first, a fixed value for elevation was used for the whole study area. To improve the accuracy of this model, the introduction of real elevation data, used in the calculation of potential evapotranspiration, could be considered. Moreover, NASA is developping a new MODIS product, MOD21, which provides the same data as MOD11 (emissivity, temperature, timepass). However, it corrects a bias known in MOD11A "over arid and semi-arid regions" (NASA, 2016). It is not available yet, but using this product instead of MOD11 could improve the results of the model, especially in dry areas. Finally, the calculation of soil net radiation (Rn_c) used a fixed extinction coefficient ($K = 0.6$), but it should be adapted to vegetation type.

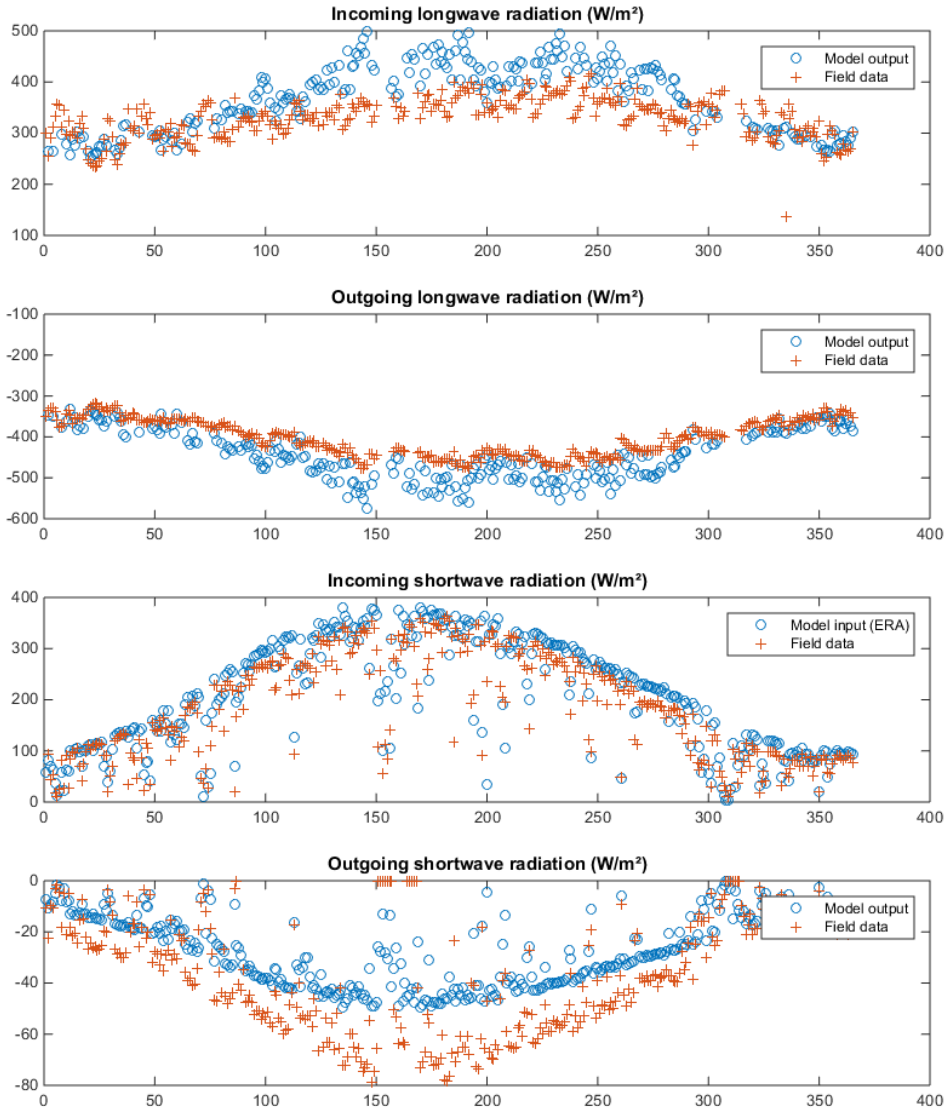

Figure 10: *Year 2011 - Graphs of 4 validation variables: LWi, LWo, SWi and SWo (in* $\mathrm{W\,m^{-2}}$*), ploted against time (in days). On each graph, model output (or input) is drawn with circles, and field data are shown with crosses*

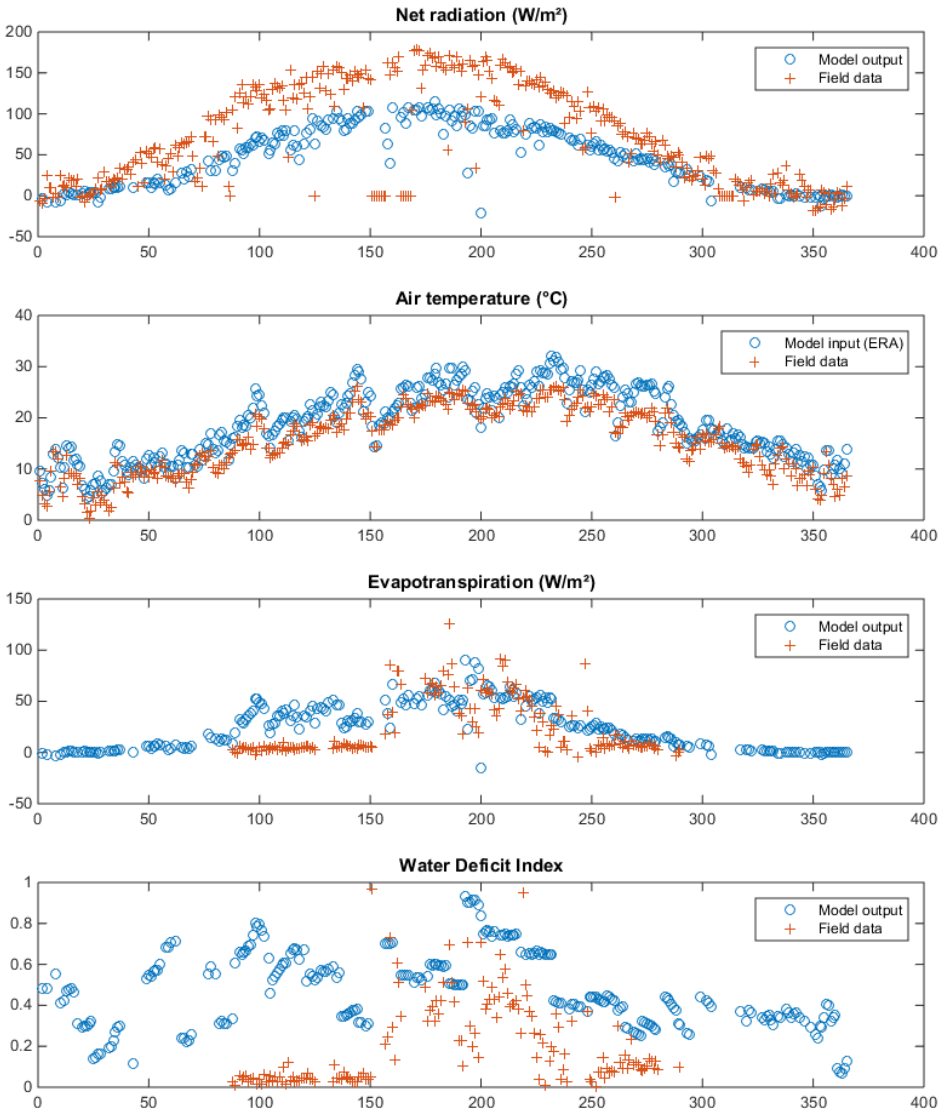

Figure 11: *Year 2011 - Graphs of 4 other validation variables: Rn ($\mathrm{W\,m^{-2}}$), air temperature (°C), evapotranspiration ($\mathrm{W\,m^{-2}}$) and ER, ploted against time (in days). On each graph, model output (or input) is drawn with circles, and field data are shown with crosses*

Table 4: *Validation results - Rn means Net radiation, LWi means Incoming longwave radiation, LWo means Outgoing longwave radiation, SWI means Incoming shortwave radiation, SWo means Outgoing shortwave radiation, ER means Evapotranspiration Ratio - Pearson's r, Root-Mean-Square Error (RMSE) and Mean Absolute Error (MAE) are calculated between model output or input in Coussoul and field validation data from Coussoul - The last column indicates the number of data points available for comparison*

Variable	Pearson's r	RMSE	MAE	Validation points (n)
Rn	0.933	48.9 $W\,m^{-2}$	41.7 $W\,m^{-2}$	248
LWi	0.385	90.5 $W\,m^{-2}$	58.8 $W\,m^{-2}$	251
LWo	0.432	90.4 $W\,m^{-2}$	50.4 $W\,m^{-2}$	251
SWi	0.944	40.4 $W\,m^{-2}$	30.1 $W\,m^{-2}$	364
SWo	0.591	25.8 $W\,m^{-2}$	13.4 $W\,m^{-2}$	365
Evapotranspiration	0.512	27.3 $W\,m^{-2}$	22.4 $W\,m^{-2}$	137
ER	0.406	0.420	0.375	137
Air temperature	0.968	3.45 °C	3.05 °C	365

III. Detection of degradation hotspots

After validation, we evaluate temporal trends. First, we look at spatial patterns in some input or output variables

To get a visual idea of the spatial patterns in Evapotranspiration Ratio (ER), we compare it to a land cover map of our study area. Figure 12 shows a comparison between land cover types from MODIS product MCD12C1 (right) and mean annual ER from our simulations (2011, left). In the south of the area, we can recognize the Mediterranean coast (water bodies are masked and appear in dark blue, as do missing data points) and the lakes in the Rhône delta, in Camargue (south-west). In addition, it seems like places with high ER (yellow pixels) correspond to the areas of mixed forest (blue-green).

Then, we compare years 2007 and 2009, because the latter seems wetter than the former: evapotranspiration and ER are globally higher in 2009 than in 2007. This seems to apply to every kind of vegetation.

In order to have an idea of the spatial variability of input variables and stress factors, we visualized maps of such parameters. For instance, Figure 13 shows mean Normalized Difference Vegetation Index (NDVI) during years 2007 (left) and 2009 (right). Figure 14 shows mean f_{SM-ATI} during years 2007 (left) and 2009 (right). Figure 15 shows mean evapotranspiration during years 2007 (left) and 2009 (right). Figure 16 shows mean Evapotranspiration Ratio (ER) during years 2007 (left) and 2009 (right).

Figure 17 is a graph of the daily average evapotranspiration over the study area, in 2007 (left) and 2009 (right). We can see that there was a plateau at 80 $W\,m^{-2}$ between days 125 and 225 in 2007, while 2009 shows more of a peak shape. Moreover, plant evapotranspiration (canopy evapotranspiration) dominates soil evapotranspiration as the main component of total evapotranspiration (LE). Since LE follows the seasonal radiation course, the ecosystem is not limited by water so much.

After studying some spatial patterns, we tried to find temporal trends in the study area. For this, we used our indicator, the Evapotranspiration Ratio (ER), and calculated the correlation between time and the annual mean of ER, for every pixel of the study area. Figure 18 shows the correlation coefficient r: values are between 1 and -1, these two extremes showing the highest correlations. Missing data appear in green (0).

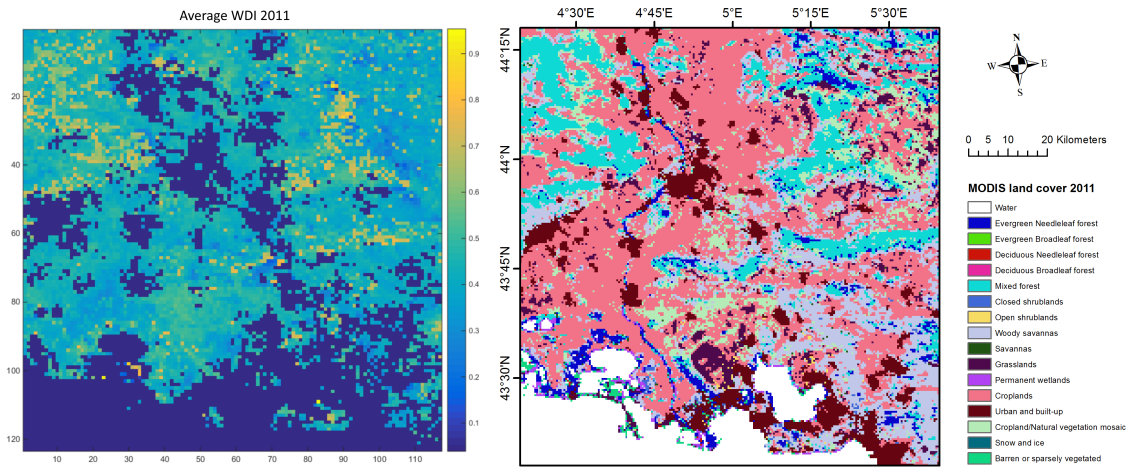

Figure 12: *Land cover from MODIS product MCD12C1 (2011) on the right, mean annual ER (2011) from the model on the left - Study area*

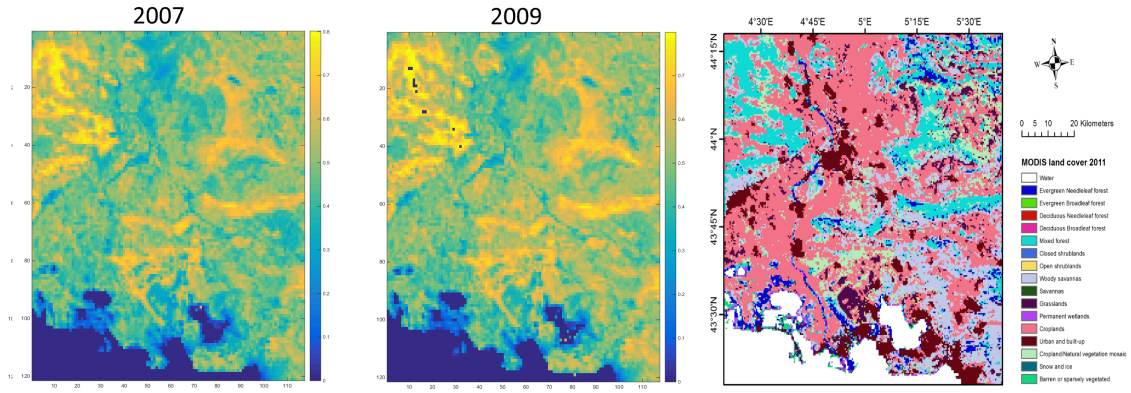

Figure 13: *Mean NDVI (Normalized Difference Vegetation Index) in years 2007 (left) and 2009 (center), land cover from MODIS product MCD12C1 (right) - Study area - Deep blue pixels in the two maps on the left indicate water bodies or missing data*

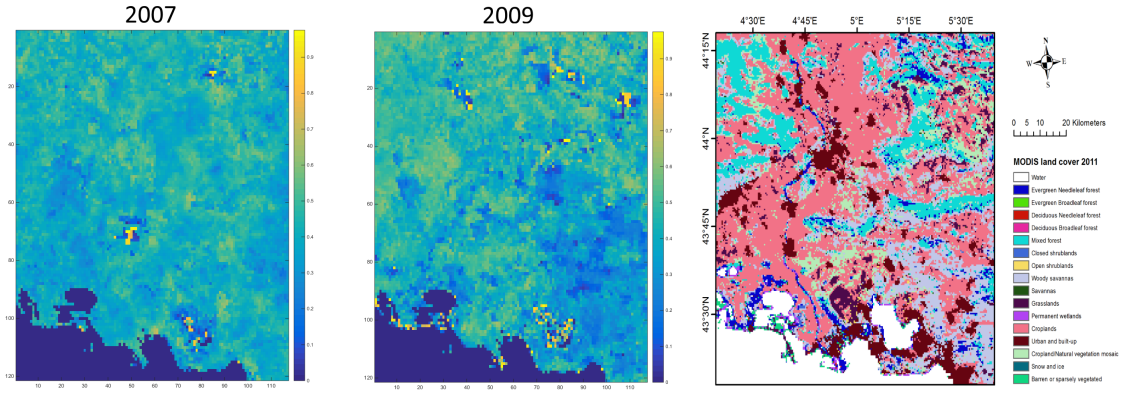

Figure 14: *Mean f_{SM-ATI} (soil moisture constraint) in years 2007 (left) and 2009 (center), land cover from MODIS product MCD12C1 (right) - Study area - Deep blue pixels in the two maps on the left indicate water bodies or missing data*

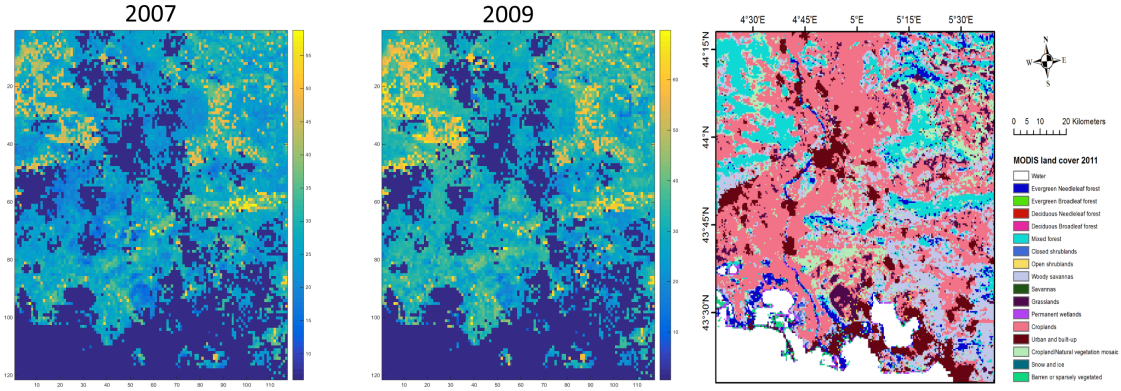

Figure 15: *Mean evapotranspiration in years 2007 (left) and 2009 (center), land cover from MODIS product MCD12C1 (right) - Study area - Deep blue pixels in the two maps on the left indicate water bodies or missing data*

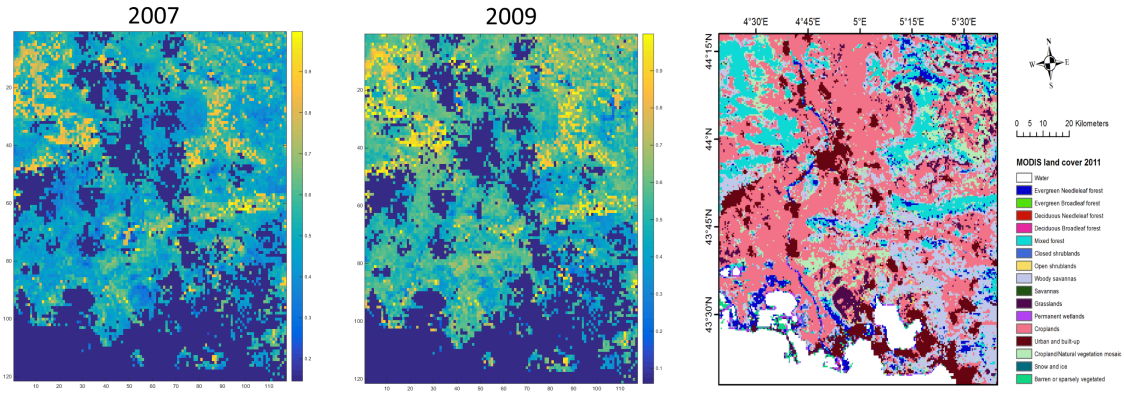

Figure 16: *Mean ER (Evapotranspiration Ratio) in years 2007 (left) and 2009 (center), land cover from MODIS product MCD12C1 (right) - Study area - Deep blue pixels in the two maps on the left indicate water bodies or missing data*

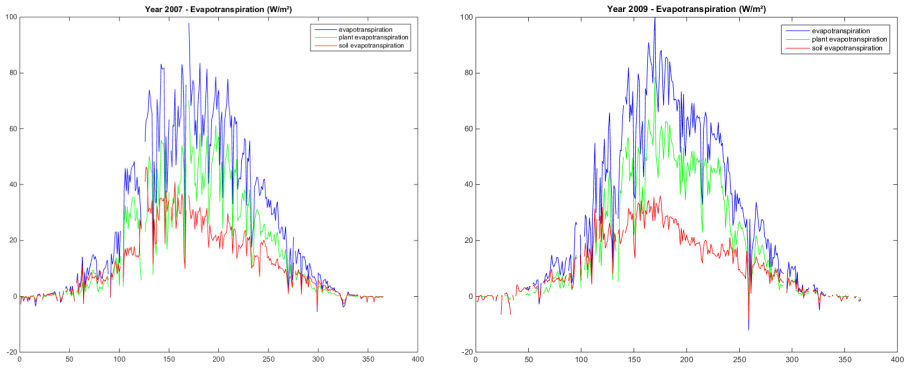

Figure 17: *Year 2007 (left) and 2009 (right) - Daily mean evapotranspiration* ($\mathrm{W\,m^{-2}}$) *against time (days) - Average over the whole study area*

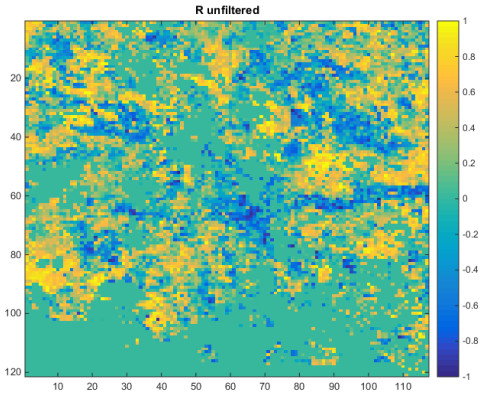

Figure 18: r^2 *unfiltered - Study area*

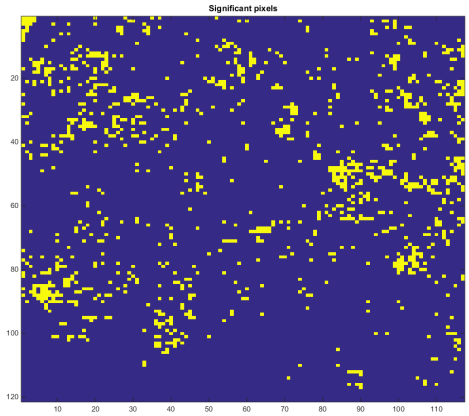

Figure 19: *Significant pixels (p-value<0.05) appear in yellow, non-significant pixels appear in blue - Study area*

Figure 19 shows the map of significant pixels: 1, in yellow, indicate that the pixel sees a significant trend in dryness change, that could be indicating degradation or restoration.

Finally, Figure 20 shows the r value for significant pixels: when the p-value was below 0.05, the r-value was set to NaN (not a number, i.e. missing data point). In this figure, yellow and orange pixels indicate high correlation between time and ER, while blue pixels indicate high negative correlation between these two variables. For instance, pixel (49,83), 43.9 °N, 5.23 °E, has a r-value of 0.94, and a p-value of 0.0013.

Figure 21 shows the increasing trend in ER in this place: the linear regression follows a slope of 0.015 /year. According to land cover data from MODIS, it is an area of cropland.

Figure 22 shows the daily values of ER in pixel (49,83), if available, for every year we considered. Even years, which we didn't study, were omitted in the time axis. The global trend is not as obvious as in Figure 21 (annual means).

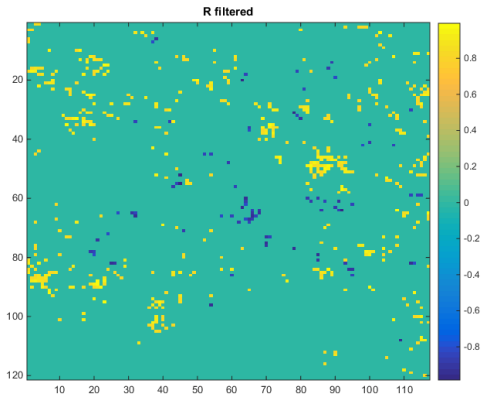

Figure 20: *r^2 filtered - pixels with low significance (p-value>0.05) were masked - Study area*

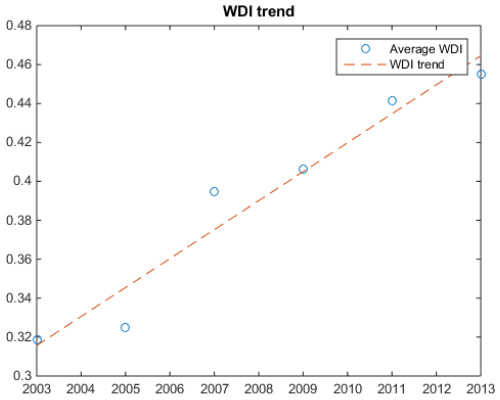

Figure 21: *Average ER trend in pixel (49,83), against time (in years). ER values (circles) and linear regression (dashed line)*

Still, we see that the maximum value is not the same each year: the maximums were lower in 2005, 2009 and 2011 than in 2003, 2007 and 2013. Finally, this detection of a potential degradation point doesn't give any information about the causes of this degradation: a closer study of the area, probably on the field, would be necessary to understand the dynamics at work there. A trend towards higher humidity could simply be related to increased precipitation.

Originally, as it is reflected in the introduction, we wanted to focus on the link between land degradation and biodiversity. We lacked time to do so, but that would be an interesting topic to investigate, provided biodiversity data can be retrieved.

VI. Conclusion

This study presents a way to detect land degradation hotspots through a spatial analysis of time trends in Evapotranspiration Ratio (ER), the ratio of evapotranspiration over potential evapotranspiration. It is based on a daily implementation of the PT-JPL model, and considers correlation between time and ER as an indicator of degradation: significant (p<0.05) pixels with r^2 over 0.8 were classified as possibly degraded.

The study area chosen for this pilot study was located in the south of France. Data came mainly from NASA's MODIS, but this study introduced ERA-Interim data to the model: air temperature and downwards surface solar radiation from this source proved to be well correlated with field data from the validation site in Coussoul. However, validation showed that albedo data from MODIS were affected by clouds and missing data points, which forced us to use a value fixed in time and space in order to obtain satisfying outgoing solar radiation values. It should be noted that flux data from Coussoul proved particularly unreliable, with an average closure error of 183%.

In addition to these findings, this study put a huge emphasis on script improvement: this included some corrections directly in the code, but also the definition of a workflow for the user. Based on a framework that can be freely downloaded, it minimizes the number of manual operations to be made by the user, so that he or she can focus on the science and not on the technical aspects of running a simulation.

Acknowledgements

The author wishes to thank Monica Garcia, Sheng Wang, Matti Kummu and Sébastien Garrigues for their guidance, help, support and valuable advice.

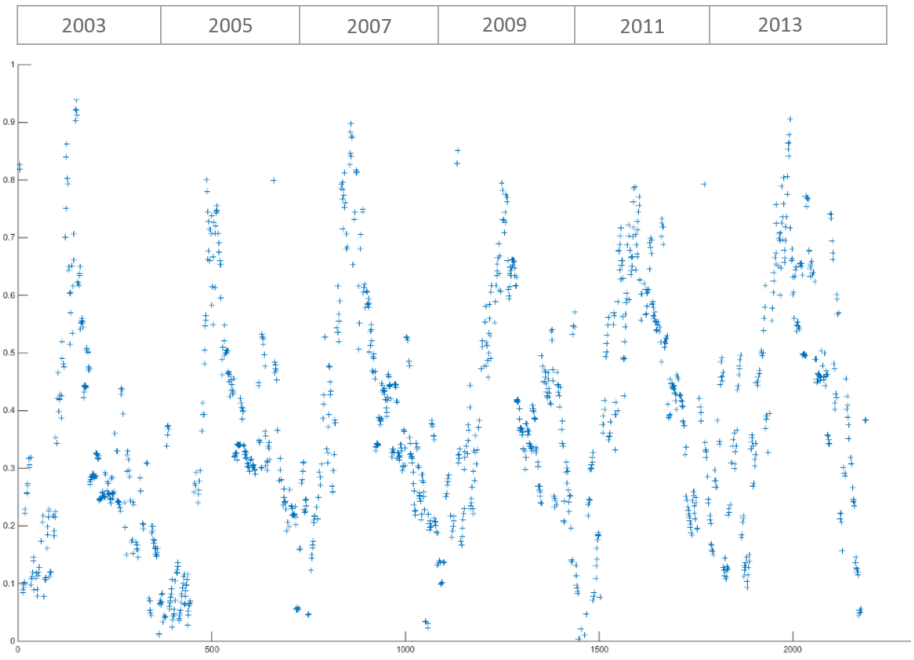

Figure 22: *Daily values of ER in pixel (49,83). Even years do not appear in the time axis.*

References

Martha C Anderson, John M Norman, John R Mecikalski, Jason A Otkin, and William P Kustas. A climatological study of evapotranspiration and moisture stress across the continental united states based on thermal remote sensing: 1. model formulation. *Journal of Geophysical Research: Atmospheres*, 112 (D10), 2007.

CBD. Strategic plan for biodiversity 2011-2020, including aichi biodiversity targets, jun 2016. https://www.cbd.int/sp/.

Dennis C Duro, Nicholas C Coops, Michael A Wulder, and Tian Han. Development of a large area biodiversity monitoring system driven by remote sensing. *Progress in Physical Geography*, 31(3):235–260, 2007.

EoE. Biodiversity?, apr 2016. https://eoearthlive.wordpress.com/view/article/150560/, temporarily not accessible.

Karin S Fassnacht, Warren B Cohen, and Thomas A Spies. Key issues in making and using satellite-based maps in ecology: A primer. *Forest Ecology and Management*, 222 (1):167–181, 2006.

Miguel Fernandez, Monica Garcia, and Nestor Fernandez. Global ecosystem restoration index. *The GEO Biodiversity Observation Network (GEO BON): Enhancing Biodiversity Observations and Products for User Needs. Twelfth Plenary Session*, pages 14–15, 2015.

Nestor Fernández. Earth observation for species diversity assessment and monitoring. *Earth observation of ecosystem services*, pages 151–177, 2013.

Joshua B Fisher, Kevin P Tu, and Dennis D Baldocchi. Global estimates of the land–atmosphere water flux based on monthly avhrr and islscp-ii data, validated at 16 fluxnet sites. *Remote Sensing of Environment*, 112(3):901–919, 2008.

Joshua B Fisher, Yadvinder Malhi, Damien Bonal, Humberto R Da Rocha, Alessandro C De Araujo, Minoru Gamo, Michael L Goulden, Takashi Hirano, Alfredo R Huete,

Hiroaki Kondo, et al. The land–atmosphere water flux in the tropics. *Global Change Biology*, 15(11):2694–2714, 2009.

Mónica García, Cecilio Oyonarte, Luis Villagarcía, Sergio Contreras, Francisco Domingo, and Juan Puigdefábregas. Monitoring land degradation risk using aster data: The non-evaporative fraction as an indicator of ecosystem function. *Remote Sensing of Environment*, 112(9):3720–3736, 2008.

Monica García, Inge Sandholt, Pietro Ceccato, Marc Ridler, Eric Mougin, Laurent Kergoat, Laura Morillas, Franck Timouk, Rasmus Fensholt, and Francisco Domingo. Actual evapotranspiration in drylands derived from in-situ and satellite data: Assessing biophysical constraints. *Remote Sensing of Environment*, 131:103–118, 2013.

GEO-BON. Essential biodiversity variables, jun 2016. http://geobon.org/essential-biodiversity-variables/ebv-classes-2/.

AR Huete and RD Jackson. Suitability of spectral indices for evaluating vegetation characteristics on arid rangelands. *Remote sensing of environment*, 23(2):213IN1–232IN8, 1987.

PG Jarvis. The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 273(927):593–610, 1976.

Jetse D Kalma, Tim R McVicar, and Matthew F McCabe. Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data. *Surveys in Geophysics*, 29(4-5):421–469, 2008.

William P Kustas and John M Norman. Evaluation of soil and vegetation heat flux predictions using a simple two-source model with radiometric temperatures for partial canopy cover. *Agricultural and Forest Meteorology*, 94 (1):13–29, 1999.

Ray Leuning, YQ Zhang, Amelie Rajaud, Helen Cleugh, and Kevin Tu. A simple surface conductance model to estimate regional evaporation using modis leaf area index and the penman-monteith equation. *Water Resources Research*, 44(10), 2008.

Maria Mira, Albert Olioso, Belen Gallego-Elvira, Dominique Courault, Sébastien Garrigues, Olivier Marloie, Olivier Hagolle, Pierre Guillevic, and Gilles Boulet. Uncertainty assessment of surface net radiation derived from landsat images. *Remote Sensing of Environment*, 175:251–270, 2016.

L Morillas, M García, H Nieto, L Villagarcia, Inge Sandholt, MP Gonzalez-Dugo, PJ Zarco-Tejada, and F Domingo. Using radiometric surface temperature for surface energy flux estimation in mediterranean drylands from a two-source perspective. *Remote Sensing of Environment*, 136:234–246, 2013a.

Laura Morillas, Ray Leuning, Luis Villagarcía, Mónica García, Penélope Serrano-Ortiz, and Francisco Domingo. Improving evapotranspiration estimates in mediterranean drylands: The role of soil evaporation. *Water Resources Research*, 49(10):6572–6586, 2013b.

Maria Carmen Moyano, Monica Garcia, Lucia Tornos, Laura Recuero, Alicia Palacios-Orueta, and Luis Juana. Assessing regional crop water demand using a satellite-based combination equation with a land surface temperature component. In *EGU General Assembly Conference Abstracts*, volume 17, page 15503, 2015.

NASA. Modis land surface temperature and emissivity (mod21), jun 2016. http://modis.gsfc.nasa.gov/data/dataprod/mod21.php.

Grey S Nearing, M Susan Moran, Russell L Scott, and Guillermo Ponce-Campos. Coupling diffusion and maximum entropy models to estimate thermal inertia. *Remote sensing of environment*, 119:222–231, 2012.

Nathalie Pettorelli, Martin Wegmann, Andrew Skidmore, Sander Mücher, Terence P Dawson, Miguel Fernandez, Richard Lucas, Michael E Schaepman, Tiejun Wang, Brian O'Connor, et al. Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. *Remote Sensing in Ecology and Conservation*, 2016.

Christopher S Potter, James T Randerson, Christopher B Field, Pamela A Matson, Peter M Vitousek, Harold A Mooney, and Steven A Klooster. Terrestrial ecosystem production: a process model based on global satellite and surface data. *Global Biogeochemical Cycles*, 7(4):811–841, 1993.

CHB Priestley. On the assessment of surface heat flux and evaporation using large scale parameters. In *Mon. Weather Rev*. Citeseer, 1972.

Stephen D Prince, De Colstoun, E Brown, and LL Kravitz. Evidence from rain-use efficiencies does not indicate extensive sahelian desertification. *Global Change Biology*, 4(4):359–374, 1998.

Woody Turner, Sacha Spector, Ned Gardiner, Matthew Fladeland, Eleanor Sterling, and Marc Steininger. Remote sensing for biodiversity science and conservation. *Trends in ecology & evolution*, 18(6):306–314, 2003.

UNCCD. *United Nations Convention to Combat Desertification in Those Countries Experiencing Serious Drought and/or Desertification, Particularly in Africa*. United Nations, 1996.

Willem W Verstraeten, Frank Veroustraete, Corné J van der Sande, Ief Grootaers, and Jan Feyen. Soil moisture retrieval using thermal inertia, determined with visible and thermal spaceborne data, validated for european forests. *Remote Sensing of Environment*, 101 (3):299–314, 2006.

WWF. How many species are we losing?, jun 2016. http://wwf.panda.org/about_our_earth/biodiversity/biodiversity/.

YQ Zhang, FHS Chiew, L Zhang, R Leuning, and HA Cleugh. Estimating catchment evaporation and runoff using modis leaf area index and the penman-monteith equation. *Water Resources Research*, 44(10), 2008.

Appendix: PT-JPL script – a tutorial

This appendix is aimed at the user who wants to run the PT-JPL daily evapotranspiration model. It explains the file structure we provide, and the workflow to be followed.

1. The lists of files

The main script takes daily binary files as input (one for each day of the year), and produces daily binary files as output. Most of the original data come as .hdf files, with variable time frequency. .hdf files need to be converted to binary, which are the input to the main script.
The main script, likes the conversion (hdf to binary) scripts, works with several lists of file. These lists contain the names of .hdf files or binary files, in chronological order.
For each variable, there is a list of the .hdf files that were downloaded. Then, there is a list of the binary files that result from the conversion from hdf to binary. Finally, in the case of data that are not produced daily, there is a second list of binary files: since the model runs daily, we need a list with 365 files. This last list contains the same files as the previous one, except these ones are duplicated according to the time step.
Let's take the Modis product MOD13A2 as an example. From the .hdf files is produced a list of their names: *Vl_T_HDF.txt*. The structure already contains a list of binary files: *NDVI_T.txt*. The h*df_to_binary_NDVI_T_tr.m* script writes these binary files, but because MOD13A2 data come every 16 days, there are only 23 of these files, not 365. *NDVI_Tin.txt* is a list of file names that artificially creates 365 input binary files: from *NDVI_T.txt*, each line is duplicated 15 times (that is, the 16 first lines are identical). The main script uses the same binary file for the 16 first days, then the second file for the 16 next days, etc.
This procedure is identical for all datasets that are not produced daily.

2. File structure

The directory "Spatial script –structure" contains all the scripts, in addition to the folders that hold input and output data.
'Root' directory:

- *spatial_script.m* : the main script
- *parameters.m* : spatial extent of the project; called by *spatial_script.m*
- *check_results.m* : visualization of some output variables; can be customized to your own needs.

'Root/Data' directory:

- Modis folders (MCD43B3, MOD11A1 etc): contain a Matlab file (*hdf_to_binary_xxx.m*) that converts Modis files (HDF format) to binary files. The text files are lists of filenames used as output by the *hdf_to_binary_xxx.m* script and/or as input by *spatial_script.m*. Data downloaded from EarthData (see later) have to be placed in these folders.
- MOD11A1: contains *eos2dump.exe*, a program that needs to be used to produce latitude data from a .hdf file (only once for a given MODIS tile).
- MOD15A2: contains *faparmax.m*, a script that needs to be run before *spatial_script.m*.
- Meteo folder: data from the ERA-Interim dataset (temperature at 2 meters and downwards surface solar radiation) have to be placed in this directory. *ERA2.m* processes these data. *NOAA_trise_tsunset.xlsx* contains data relative to time of sunrise and sunset.

- Input_files_generator.m creates lists of files names; these files have already been generated, but this script can be useful in the case of a leap year (said files then lack one line).

'Root/Results' directory:

- Each directory contains lists of output filenames. Output from *spatial_script.m* will appear in these folders.

File structure

3. Workflow (these steps are detailed below):

1) Create the folder structure (copy-paste "Spatial script – structure")
2) Download MODIS data, and put the HDF files in the adequate folder (inside the "Data" folder)
3) Dump latitude (eos2dump, needs to be done only once for a given MODIS tile)
4) Create the files lists (command prompt 'dir')
5) Run *hdf2binary.m* (Matlab)
6) Download ERA data.
7) Run *ERA2.m* (Matlab)
8) Run *spatial_script.m* (Matlab)

2) Data download (HDF files)

https://search.earthdata.nasa.gov/

This website doesn't work with Chrome anymore. Firefox recommended, with the plugin "DownloadThemAll" (DtA).

Search by product name (MOD11A2…), set the time frame (from January 1st to December 31st). Regarding the spatial limits, several methods are available. If one is looking for a single tile, the easiest way is to point at the map.

Select the collection on the left panel, then click "Retrieve collection data". Select data access via "Download", then "View download links".

Right-click in the page, then launch the DtA plugin. Ctrl+A to select all links, right-click "Tick all selected", then start the download.

Attention: when setting the time frame, some files may be missing. The lists of filenames (those with daily entries) need to be updated accordingly: missing days are assigned data from the last available date. Sometimes, EarthData shows more files than necessary (usually from the year before). Don't select these files with DtA, in order to download only the necessary files.

3) HDF-EOS2 DUMP

This operation extracts latitude data from a .hdf file.

'Data/MOD11A1' contains the program *eos2dump.exe* (Source, with examples). In Windows' command prompt, run the following (adapt the .hdf file name to one of those present in your folder):

```
$eos2dump -c1 MOD11A1.A2007113.h11v09.005.2007136163924.hdf > lat_MOD11A1
```

A binary file *lat_MOD11A1* is created, it can be used in any simulation taking place on the current MODIS tile. Therefore, this operation needs to be completed only once, provided you keep working on the same area.

4) and 5) Data conversion (HDF to binary)

First a list of .hdf file names must be created. For this, use the Windows command prompt, with the following command: dir *.hdf/b > xxx_HDF.txt (see the section "file names" below to name the list of .hdf files).

This creates a text file containing the list of all HDF files present in the folder you're working in.

The lists of binary files are already in the structure, so the *hdf2binary.m* script can be run now. It calls all the conversion scripts, and *faparmax.m*.

6) and 7) ERA-Interim data

Temperature and incoming short wave radiation data do not come from a Modis product but from a meteorological reanalysis dataset called ERA-Interim. Download from the following page:
http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/
Two files need to be downloaded and stored in the folder 'Root/Data/meteo', with the following parameters:
Time=12, step=0, 2 metre temperature (2nd column, 1st line). Call this file *ERA_t2m.nc*
Time=0, step=12, Surface solar radiation downwards (2nd column, 35th line). Call this file *ERA_rad.nc*
Then run *ERA2.m*

ARCHIVE, no longer necessary (except in case of leap years: the defaults files contain only 365 lines)

Creating the input/output files list

This is done with the *output_file_generator.m* script (or *input_file_generator.m)* located in 'Root/Results'.

```
year =[1:1:365]; %take care of leap years

format ='WDI%1.0f\n';
fileID=fopen('WDI.txt', 'w');
fprintf(fileID,format, year);
fclose(fileID);
```

The names need to be adapted to reflect on what the output is. In this case, we calculate WDI. The script creates a text file with 365 different lines, from "WDI1" to "WDI365". These are the names of the 365 output files. For input files, the number of lines depends on the data: 23 lines if they are produced every 16 days, 46 lines if they are produced every 8 days.

Duplicate file names

The result of this step is a text file containing 365 (or 366) binary file names.
This is done with a Linux Terminal.
Run the following: awk '{for(i=0;i<8;i++)print}' xxx.txt > xxxin.txt
The figure here (8) depends on the dataset frequency: 8 for data available every 8 days, 16 for data available every 16 days.
The resulting daily file will be the input for the spatial script.
Attention: the result of the command will be a file longer than 365 lines (because 365 is not divisible by 8). Check the file, and delete the unnecessary lines (3 of them, 2 in case of a leap year).
Tutorial for installing a Linux virtual machine on a Windows computer]

4. Input data from MODIS

MODIS product	Time step	Variable	Scripts
MOD13A2	16 days	1 km NDVI	**a**
MOD15A2	8 days	1 km fPAR	**b**
		1 km LAI	**c**
MCD43B3	8 days	1 km Black/White sky albedo	**d**
MOD/MYD 11A1	1 day	1 km Band 31/32 Emissivity	**e, l**
		1 km Daytime land surface temperature	**f, m**

		1 km Nighttime land surface temperature	**h, o**
		1 km Day observation time	**g, n**
MOD/MYD 11A2	8 days	1 km Band 31/32 Emissivity	**i, p**
		1 km Daytime land surface temperature	**j, q**
		1 km Nighttime land surface temperature	**k, r**

5. File names

The first level of indentation is the list of .hdf files. The second one consists of the output of *hdf_to_binary.m* scripts. The third level consists of duplicated files, for datasets that are not produced daily.

MOD13A2 (16 days):
VI_T_HDF.txt
a NDVI_T.txt
NDVI_Tin.txt

MOD15A2 (8 days):
LAIfAPAR_HDF.txt
b fAPAR.txt
fAPARin.txt
--
c LAI.tx
LAIin.txt

MCD43B3 (8 days):
Albedo_HDF.txt
d Albedo_BSA.txt
BSAin.txt
Albedo_WSA.txt
WSAin.txt

MOD11A1 (daily):
MOD11A1_HDF.txt
e Emis31_dT.txt
Emis32_dT.txt
--
f Tday_T.txt
--
g Tday_timepass_T.txt
--
h Tnight_T.txt

MOD11A2 (8 days):
MOD11A2_HDF.txt
i Emis31_T.txt
Emis31_Tin.txt
Emis32_T.txt
Emis32_Tin.txt
--
j Tday_T.txt
Tday_Tin.txt
--
k Tnight_T.txt
Tnight_Tin.txt

MYD11A1 (daily):
MYD11A1_HDF.txt
l Emis31_dA.txt
Emis32_dA.txt
--
m Tday_A.txt
--
n Tday_timepass_A.txt
--
o Tnight_A.txt

MYD11A2 (8 days):
MYD11A2_HDF.txt
p Emis31_A.txt
Emis31_Ain.txt
Emis32_A.txt
Emis32_Ain.txt
--
q Tday_A.txt
Tday_Ain.txt
--
r Tnight_A.txt
Tnight_Ain.txt

6. Variables and variable types

Input variables				**Output variables**		
Variable code	name	type		Variable code	name	type
A1	NDVI	Int16		B30	NDVI	Double
A2	fAPAR	Uint8		B31	fAPAR	Double
A3	Emis31 daily A	Uint8		B32	fc	double
A4	Emis32 daily A	Uint8		B33	fg	Double
A5	Emis31 daily T	Uint8		B34	fM	double
A6	Emis32 daily T	Uint8		B35	fT	Double
A7	Tday A	Uint16		B36	Tday A	Double
A8	Tday T	Uint16		B37	Long wave in	Double
A9	Timepass A	Uint8		B38	Long wave out	Double
A10	Timepass T	Uint8		(B39)	Net short wave daily	Double
A11	BSA	Uint16		B40	Net short wave instant	Double
A12	WSA	Uint16		B41	Net radiation instant	Double
(ERA)	Tair	Double		B42	Net radiation daily	Double
(ERA)	Short wave in inst	Double		B421	Short wave out	double
A13	LAI	Uint8		B44	Net soil radiation	Double
A14	Net radiation daily	Double		B45	Net canopy radiation	Double
A15	Fc	Double		B46	Etpc	Double
A16	Fg	Double		B47	Etps	Double
A17	fT	Double		B48	Etp	Double
A18	fM	Double		B49	Available energy	Double
				B50	Canopy latent heat	double
A19	BSA	Uint16		B51	ATI	Double
A20	WSA	Uint16				
A21	Tday T	Uint16				
A22	Tday A	Uint16				
A23	Tnight T	Uint16				
A24	Tnight A	Uint16				
A25	ATI	Double		B52	fSM ATI	Double
A26	Etps	Double		B53	Soil latent heat	Double
A27	Canopy latent heat	Double		B54	Latent heat	Double
A28	Available energy	Double		B55	Evaporative fraction	Double
A29	Etp	Double		B56	Water deficit index	Double

T = Terra satellite (MOD)

A = Aqua satellite (MYD)

(ERA) = data from the ERA-interim dataset

Output variables used as input later in the script